TECHNO:PHIL BAND 9

Jan-Hendrik Heinrichs / Birgit Beck / Orsolya Friedrich (Eds.)

Neuro-ProsthEthics

Ethical Implications of Applied Situated Cognition

1682 J.B. METZLER

Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie

Band 9

Diese Reihe befasst sich mit der philosophischen Analyse und Evaluation von Technik und von Formen der Technikbegeisterung oder -ablehnung. Sie nimmt einerseits konzeptionelle und ethische Herausforderungen in den Blick, die an die Technikphilosophie herangetragen werden. Andererseits werden kritische Impulse aus der Technikphilosophie an die Technologie- und Ingenieurswissenschaften sowie an die lebensweltliche Praxis zurückgegeben. So leistet diese Reihe einen substantiellen Beitrag zur inner- und außerakademischen Diskussion über zunehmend technisierte Gesellschafts- und Lebensformen.
Die Bände der Reihe erscheinen in deutscher oder englischer Sprache.

This book series focuses on the philosophical analysis and evaluation of technology and on forms of enthusiasm for or rejection of technology. On the one hand, it examines conceptual and ethical challenges that philosophy of technology has to face. On the other hand, critical impulses from philosophy of technology are returned to the technology and engineering sciences as well as to everyday practice. Thus, this book series makes a substantial contribution to the academic and transdisciplinary discussion about increasingly technologized forms of society and life.
The volumes of the book series are published in German and English.

Jan-Hendrik Heinrichs · Birgit Beck · Orsolya Friedrich
Editors

Neuro-ProsthEthics

Ethical Implications of Applied Situated Cognition

J.B. METZLER

Editors
Jan-Hendrik Heinrichs
Ethik in den Neurowissenschaften
Forschungszentrum Jülich GmbH
Jülich, Nordrhein-Westfalen, Germany

Birgit Beck
FG Ethik und Technikphilosophie
Technische Universität Berlin
Berlin, Germany

Orsolya Friedrich
Fakultät KSW, Institut für Philosophie
FernUniversität Hagen
Hagen, Nordrhein-Westfalen, Germany

ISSN 2524-5902 ISSN 2524-5910 (electronic)
Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie
ISBN 978-3-662-68361-3 ISBN 978-3-662-68362-0 (eBook)
https://doi.org/10.1007/978-3-662-68362-0

This J.B. Metzler imprint is published by the registered company Springer-Verlag GmbH, DE, part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Paper in this product is recyclable.

Contents

Introduction

Jan-Hendrik Heinrichs

Human cognition and emotion are deeply influenced by factors outside the human brain. We perceive by bodily action – changing visual perspective, touching, grasping, sniffing or inhaling deeply. We bodily manage emotions – cringing, posturing, crying, etc. We employ cultural techniques and engrams – letters, numerals, algorithms, special clothes, gestures rituals and places for grieving, joy, aggression. We use external mnemonic device from stone tablets to computational tablets, we structure our knowledge and communications in paragraphs, articles, tables, we calculate with tools like pen and paper, the abacus, or digital computers. And for the most part, these external factors – once included in our cognitive activities – are valued highly, or at least more highly than materially equivalent but cognitively neutral objects. A slab of stone gains importance once it is inscribed, the knotted handkerchief reminding us of a task it more important than a similarly knotted piece of cloth. Not to speak of the difference between a personalized smartphone and the same device with default factory settings.

One would suspect that these influencing factors have always played a deep and continuous role in how we think about cognition. They have not. The means of cognition, that is, the external structures, tools, and scaffolds with which we think and feel found little attention in the discussions of philosophy of mind or ethics for most of the time. While the human body, especially the hands have received some attention in explanations of cognition, body-external tools go widely unnoticed with a few notable exceptions. This was particularly true during the so-called decade of the brain 1990–2000. The marginalization of everything beyond the brain

J.-H. Heinrichs (✉)
Institute for Neuroscience and Medicine 7: Brain and Behaviour, Forschungszentrum Jülich, Jülich, Germany
e-mail: j.heinrichs@fz-juelich.de

J.-H. Heinrichs et al. (eds.), *Neuro-ProsthEthics*, Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie 9,
https://doi.org/10.1007/978-3-662-68362-0_1

was in part a consequence of a tacit theoretical stance in the neurosciences, which has been dubbed intracranialism.

Intracranialism is either the substantial claim that everything mental goes on in the brain, or the epistemic claim that all we need in order to explain human behaviour are information about the brain (see Adams & Aizawa, 2008). While intracranialism has ancient precursors reaching back to Hippocrates (in his On the sacred disease), its modern version owes to the success and to the enthusiastic reception of neuroscientific research since the 1990s.

In philosophy the decade of the brain and the prior and consecutive years were characterized primarily by intense debates about reductionisms and eliminativisms implicit in the intracranialist thesis. Eliminativist and reductionist approaches were not new, quite the opposite, they had been common fare in philosophy of mind at least since the heydays of logical positivism. But earlier reductionisms had rarely ever taken into account real results from the neurosciences. They were based on speculative, neurosciences, imagined to use explanatory models from basic physics, which then still dominated philosophy of science. This changes quite significantly in the decade of the brain, maybe even slightly before with Patricia Churchlands Neurophilosophy (Churchland, 1986). More recent debates therefore focussed on the question, how neuroscientific explanation – often in contrast to physical explanation – proceeds and whether it suffices to explain human cognition (Bickle, 2003; Craver, 2009). This was interesting progress in the philosophy of mind, but it did not touch upon the common neglect of external factors of cognition in the discipline. More surprisingly, neither did most of the early anti-reductionist and anti-eliminativist answers. They did insist that even a full physical description of a cognizing organism would not suffice to explain its cognition; several did develop externalisms of meaning or content, making the latter depend on an organisms' history in its environment (Davidson, 1987; Putnam, 1975). The task to point out that external factors do not just contribute to the content, but to the processes of cognition themselves was left to later externalist theories.

Ethical considerations of the implications of both intracranialism and its opposition, tended to be in the background of the debate. The central, broadly ethical themes of that time were free will on the one hand and personhood and personality on the other. Both were a reaction to claims about the allegedly novel neuro-determinism, i.e., the thesis that events in the brain fully determine our mental properties and processes. For many of these debates it did, however, not play an important role what the determining forces were, main contributions focused on determinism in general and much less on *neuro*-determinism. Rather they prolonged a debate which had been ongoing when philosophy of mind still took explanation in physics as its paradigm case.

This debate was one of the few taking into account external factors of thinking early on, if in a version widely detached from real world neurotechnologies, namely in the form of fictional neurotechnologies figuring in a series of thought experiments. One of the most famous figures in the free will debate is the mad neuroscientist and his mind-control apparatus (Frankfurt, 1971), closely followed by the vat, which is able to hold a brain alive and fully stimulated with real live

equivalent perceptions (Putnam, 1981) and the experience machine, which basically does the same thing, but without detaching the body first (Nozick, 1974). While these have not served to exemplify real effects of neurotechnologies, they nevertheless shaped the debate, setting the threshold for interesting neurotechnologies quite high.

The debates about neuro-determinism gave voice to calls for reforms of educational and penal practice on the basis of supposed results of neuroscience (Haidt, 2001). These calls tended to be rather simplified inferences from deterministic interpretations of neuroscientific results and were rarely systematic analysis in legal or educational ethics. More detailed analyses have been engaged in slightly later, if often still based on strong interpretations of neuroscientific results (for example Greene & Cohen, 2006). In neither did the external factors of cognition, its embodiment and its scaffolds and tools play any relevant role.

Real world neuro-prosthetics on the other hand, have been considered in depth, but exclusively in bio-ethical discussions. They have primarily undergone close scrutiny because of their invasiveness and possible distorting influence on intracranially realized decision-making processes (Klaming & Haselager, 2013). Following established, if often tacit bioethical assumption, the more invasive techniques came under more intense investigation.

In particular, deep brain stimulation (Emily et al., 2009; Mashour et al., 2005; Schlaepfer & Fins, 2010) and interfaces between humans and computational or robotic components (Jebari, 2013; Soekadar et al., 2008) were the focus of bioethics very early on. They were, however, not discussed with regard for the recent progress in the philosophy of mind, but rather with a focus on the practical implications of possible – and near future– interventions. The exception to this is the discussion about the influence of stimulation procedures on the freedom of decision of humans (Gilbert, 2015; Kraemer, 2013; Roskies, 2006). Even these contributions were primarily conducted from a bioethical perspective and less from a firm stance in the philosophy of mind debate about free will. This might – amongst others – have to do with the fact that the debate about free will used the above-mentioned speculative technologies as intuition pumps and refrained from grappling with the effects of real world technologies.

The central topic of many discussions of neuro-prosthetic and stimulatory devices was the potential threat to individual autonomy and qualitative personal identity. Only few contributions pointed out that neuro-stimulatory devices can potentially counter or mitigate the much larger threat to autonomy and personal identity posed by neurodegenerative diseases and thus uphold a person's decision-making capacity and personality (Synofzik & Schlaepfer, 2008). Contributions discussing this supportive effect of neuro-prosthetic and neuro-stimulatory devices predominantly focused on therapeutic settings. In addition, there is a broad discussion about human enhancement, i.e. the improvement of healthy cognition beyond some norm (Heinrichs et al., 2022). Given the extremely limited real-world examples of successful enhancement, this discussion had, however, to remain mostly speculative. Consequently, the ethics of neuro-prosthetic and stimulatory devices saw the latter as either a threat to cognition or as a means to compensate for losses

of cognitive ability – in extreme cases to improve cognitive ability, but not as a means to shape cognition.

In a nutshell, tools of cognition played little role in the mainstream of philosophy of mind and of cognitive science and were seen as either a therapeutic device or a threat to cognition in ethics. As mentioned above, this does not square with their ubiquity in human cognition. Neither does it fit the importance such tools are usually assigned in individual as well as collective valuation.

This neglect ended with the advent of two separate traditions in philosophy: on the one hand praxeology in the philosophy of science and on the other situated cognition theory. With 'praxeology' in the philosophy of science we refer to investigations into the cognitive processes as realized in the laboratory, shaped by the social dynamics of institutes and research groups, as well as by the diverse types of scientific equipment (Knorr-Cetina, 1999). This type of investigation into the practice of science (Giere, 2002; Nersessian, 2009) highlighted how even the most rigorous cognitive processes, those in scientific research do not simply follow some rationalist ideal, but are indeed shaped by social and especially by artefactual contexts (Heersmink, 2016). In fact, it had direct repercussions for the understanding of understanding and of *Erkenntnis*, both in philosophy of science and in philosophy of mind. Both are not sufficiently characterized by methodological and alethic criteria, but result from processes of design and production of technological, social, and methodological norms and contexts. At the same time, philosophy of scientific practice opened a perspective on the close interplay between control over artefacts, social dynamics and the trajectory of scientific research and its cognitive processes. As such, the praxeological view always encompasses the social contexts of cognition, and thus allows for easy connections to ethical investigations.

The second tradition, situated cognition theories, takes a broader field of cognitive processes into view, showing how not only scientific but all cognition is embodied, socially and environmentally embedded and possibly extended by tools and technologies. Situated cognition, or 4E cognition, is a form of externalism with regard to the brain (Robbins & Aydede, 2009). It is a form of active externalism. That is, it does not merely bind the meaning or content of a certain thought to the thinker's history or environment, but claims that the very cognitive state or process depends on or is even co-constituted by something external to the thinking brain. In the case of Embodiment, this external contributor is the thinker's body, in the case of Embeddedness or Extension it is their environment which shapes or even co-constitutes cognition, and for Enactivism it is the thinker's action. The different approaches under the 4E umbrella are united by their opposition to the intracranialism described above. They insist that it is impossible to derive a full theory of cognition without taking into account information about processes outside the brain. Depending on the exact version of a situated cognition approach they go beyond this epistemic claim and take cognition to be a process that is not constrained to the brain. Several explanatory projects in cognitive science have taken up one or the other version of situated cognition approach and generated novel and powerful explanations (Newen et al., 2018; Shapiro, 2014).

In addition to its impressive impact in the philosophy of mind and the sciences both 4E cognition and the praxeological method quickly made an impression in ethics. Many of the early seminal contributions to situated cognition theories such as Varela, Thompson, and Rosch's *The embodied mind* (Varela et al., 1991), Clark and Chalmers *The extended mind* (Clark & Chalmers, 1998), and even Hutchins *Cognition in the wild* (Hutchins, 1995) contained explicit ethical considerations on the moral significance of their insights. One of the main drivers of this impression on ethics was the subdiscipline of neuroethics and in particular Neil Levy's book with the same title (Levy, 2007). Levy suggested to base neuroethics on the extended mind theory and stated that the moral reasons for or against interventions into external realizers of cognition are morally on par with interventions into biological realizers, i.e., the brain. The parity principle and related methods rendered both neuro-prosthetics and external tools of thinking morally significant in the same way. This idea took root in and beyond neuroethics and resulted in a growing field of literature on the practical and moral dimensions of cognitive and emotive tools.

However, there was one issue that arose in general ethical theorising, which did not much bother neuroethical thought: the issue of delimitating the scope of morally relevant tools of thinking. Neuroethics had to take relatively little account of this issue, because its scope was fixed by different sets of considerations. Beyond that, however, it would suddenly seem that *all* tools of thinking come to be relevant for ethical evaluation. This, in turn, burdened the ethics of cognitive tools with a problem similar to the main objection against situated cognition approaches in philosophy of mind: a bloat objection. The original cognitive bloat objection in the philosophy of mind claims that situated cognition, especially extended mind theory cannot distinguish between external co-realizers of cognition and mere cognitive tools and thus extends cognitive systems so far as to make the concept useless (Adams & Aizawa, 2008). The parallel argument for the ethics of cognitive tools claims that with a situated cognition approach, all tools of cognition – not just neuro-prosthetics and neurostimulation – become morally relevant, levelling important distinctions previously employed in medical ethics and bioethics (Heinrichs, 2021).

Solutions to the bloat problem – in philosophy of mind as well as in ethics – have, however, been on offer from early on. Clark and Chalmers in their seminar article had suggested criteria to distinguish between mere tools and cognitive extenders, and several authors have refined these criteria into dimensions of integration in the aftermath (Heersmink, 2015; Heinrichs, 2018). In addition, the ethical bloat argument – other than the cognitive bloat version – has an inherent weakness. Even if – according to the ethical parity thesis – interventions into external cognitive tools become morally relevant, this does not imply that all relevant distinctions between external tools and prosthetics are levelled. While both may be morally relevant as contributors to a person's cognitive states, there are still morally relevant distinctions (Heinrichs, 2021), for example, that one of them is invasive and the other is not. An extension of the realm of ethically relevant phenomena does not imply levelling distinctions within this realm.

That implanted devices can raise issues beyond those raised by their character as tools of cognition has become particularly clear in recent years, when the first neuro-prosthetic implants became obsolete or ceased to receive support by their producers. This is a common occurrence for external tools of thinking and does not raise serious moral issues for these. For implanted devices, however, this issue is serious and has sparked a debate about long term duties for producers and adequate precautions in the healthcare system (see the discussion on Vold / Liao in this volume).

The inclusion of cognitive tools in the realm of ethically relevant phenomena went hand in hand with a change in the perception of external influences on cognition. While beforehand – informed by the model of the autonomy-threatening neuro-prosthetic or stimulatory device – these were predominantly perceived as threats to cognitive integrity, they now came to be perceived as predominantly beneficial. This effect is owed not the least to the choice of examples in seminal articles of situated cognition theories. Clark and Chalmers for example focus on the benign example of a notebook, Clark's further elaboration in his *Natural born cyborgs (*Clark, 2003*)* concentrates on useful tools, Anderson's early discussion of neuro-prosthetics from an extended mind perspective (Anderson, 2008) exemplifies the argument with a hearing aid. In recent discussion artificially intelligent devices supporting or embedding cognition have received intensified attention (Hernández-Orallo & Vold, 2019) under the speaking title of AI extenders. And while the latter raise ethical issues of their own, they were – in this debate[1] – perceived to be predominantly beneficial and means of improving human cognition.

This perception of tools of cognition as predominantly beneficial owes to an understanding of the external contributors to our cognition, which has been tacitly presupposed in this very introduction up to this point. Referring to them as cognitive tools implies a certain model of the relation between human agents and their environment, namely a model of an autonomous agent rationally choosing instruments for their existing purposes. This model might be adequate for many cases of human beings making use of their environment for cognitive purposes, but as more recent discussions in situated cognition theory show, it is not universally adequate.

Recent contributions to the debate about situated cognition approaches have begun to highlight that many of the tools and scaffolds of our cognition either do not fit a user-tool model, because they are not autonomously and intentionally chosen and employed, or while fitting the model still are not as beneficial as one might assume (Liao & Huebner, 2021; Slaby, 2016). Current and upcoming approaches try to cover these hostile scaffolds (Timms & Spurrett, 2023) in terms of situated cognition theory or complement situated cognition with another theoretical perspective. There is for example a powerful theoretical synergy between situated cognition approaches and a recent strand of theorizing in philosophy of biology namely theories of niche construction (Coninx, 2023; Nogueira de

[1] This is not withstanding the debate about AI's potential to result in forms of deskilling, including moral deskilling as in Vallor (2015).

Carvalho & Krueger, 2023). These theories provide a framework for explaining how organisms are not merely shaped by, but in turn modify their ecological niche (Laland et al., 2000). Niche construction theory can focus on long term phylogenetic processes as well as on ontogenetic niches, i.e., the ecological conditions for the development of the current and next generation of a species (Odling-Smee & Laland, 2011). When focusing on the cognitive niche, this ontogenetic version takes into account a broader set of influencing factors of cognition than situated cognition approaches and without the user-resource-model. This is particularly suited to describe developmental processes and external, heteronomous effects on cognition, but it might lack the theoretical resources to distinguish between such heteronomous influences and tools for which the user-resource model is adequate. This debate is relatively young and the relation, whether complementary, competitive, or something else, between situated cognition approaches and niche construction theories is not yet settled.

The present volume integrates the different strands in the debate about the external conditions and tools of cognition. It brings together discussion about the theoretical background of the ethics of cognitive tools and scaffolds, both from an affirmative and a critical perspective, and state of the art testcases for situated cognition approaches such as novel neurotechnologies and artificially intelligent cognitive scaffolds. It highlights the commonalities as well as the differences between neuro-prosthetic devices and other external tools or scaffolds of cognition.

Robert W Clowes, Paul Smart and Richard Heersmink discuss the perspective of extended mind theorists, who argue that non-biological, external resources like notebooks or smartphones play a crucial, constitutive role in cognitive processes. These resources are considered as potential components of the cognitive and mental machinery responsible for realizing cognitive states and functions. Within this context, the chapter delves into three areas of ethical concern related to the extended mind: mental privacy, mental manipulation, and agency. Additionally, the ethics of the extended mind is examined from the viewpoint of three broad normative frameworks: consequentialism, deontology, and virtue ethics.

Karina Vold and Xinyuan Liao take up the cause of notable instances where users of neuro-prosthetics have suddenly lost access to their sophisticated tools, revealing the vulnerable and precarious nature of these technologies. They see these instances as challenging the notion that users can consistently maintain the necessary relationship with their tools to meet the criteria of parity supporting the extended cognition theory. Particularly, these technologies appear to violate a condition of ownership that has been crucial in the literature on extended cognition for the past two decades. In this context, the paper not only argues for the inclusion of neuro-prosthetics as part of one's extended cognitive system despite challenges to the ownership condition, but also reviews the history and evaluates the current status of this ownership criterion in the literature on extended cognition. The paper contends that the ownership condition has several shortcomings and proposes introducing the concept of "co-ownership" as a necessary distinction that better explains the functioning of advanced cognitive technologies.

Jan-Hendrik Heinrichs tries to dispel the notion that cognitive tools are universally beneficial and deserving of deep integration into cognitive systems. His article advocates for a taxonomy of cognitive systems and their components that acknowledges the existence of hostile and detrimental tools and recognizes that many tools may be more suitable remaining on the periphery of an extended cognitive system. It explores the potential moral issues arising when a cognitive tool becomes deeply integrated into a cognitive system and highlights three ways in which such integration can be detrimental: narrows, detours, and dead ends. These adverse effects can pose moral challenges and necessitate careful consideration during technology development. The author suggests that this list of detrimental effects is open-ended, as further investigations and advancements in technology may reveal additional impacts on a system's cognition.

Mary Walker & Robert Sparrow draw doubt on the Extended Mind thesis by presenting an analogous argument, termed the "extended body thesis" (EBT). EBT proposes that bodily processes are not solely contained within the boundaries of the physical body but also extend into the external environment. This extension incorporates objects like machines that support respiration, circulation, mobility, or object manipulation, which can be regarded as components of a person's "extended body" when they meet certain criteria. Walker and Sparrow point out that the conclusion that bodies can be considered extended entities may initially appear counterintuitive, and it carries profound practical and ethical implications. The authors contend that proponents of the extended mind thesis must address the validity of arguments supporting EMT while potentially dismissing the analogous argument for EBT, or alternatively, they must embrace the idea that both minds and bodies can be extended. If they choose the latter option, they are obligated to provide an account of how individuals should interact with one another as entities who are both mentally and physically "in the world", entangled in various radical possibilities.

Charles Rathkopf takes issue with the conception of a brain-computer interface adequately decoding its user's intentions. This conception is taken to play a major role in determining the user's culpability. Rathkopf contends that this requirement is muddled and proposes an alternative perspective. The argument suggests that, for the purpose of evaluating moral culpability, actions facilitated by a brain-computer interface should be treated similarly to actions facilitated by regular (albeit complex) tools.

Inken Titz seeks to challenge the prevailing cognitivist bias in discussions about AI-based and general moral enhancement interventions by introducing the concept of "moral identity" as a compelling alternative. The primary aim is to establish moral identity as a significant and empirically sound focal point for moral enhancement. Unlike the conventional emphasis on cognition for preserving autonomy, the chapter argues that cognition does not occupy the central role we desire in moral conduct, including higher-order abilities. This focus on identity aligns with philosophical perspectives that consider identity as fundamental to moral agency and acknowledges that our will, rather than cognition, often influences our moral behaviour.

Stephen Rainey employs language as an example of situated cognition and investigates whether tracing back to a point where an antecedent action was taken for which there was relevant control can account for ascriptions of responsibility in cases of neuro-prosthesis mediated speech. He argues that relying on tracing to account for ascriptions of responsibility in cases of neuro-prosthesis mediated speech is not sufficient. Hence, the situation of speech facilitated by neuro-prosthetics appears to be a compelling case that should prompt examination from a situated cognition perspective. This perspective relies less on the idea of transferring mental content to external vocalization and more on complex interpersonal contexts, considering the specifics of the physical and conventional environment. A rational relations account of responsibility might be more suitable than tracing in this scenario, as it specifically addresses the content of speech, which is critical when analysing responsibility for speech. However, if content is not the central concern, tracing could still offer a rough way to attribute responsibility for the outcomes of speech.

Marco Stier formulates and tackles the following dilemma for the use of *clinical decision support systems*: it can either choose not to use AI, potentially sacrificing the best care for patients, or embrace its use, which may lead to insurmountable challenges in properly attributing responsibility. In order to tackle this dilemma, he suggests viewing the physician and the CDSS as a coupled cognitive system, as conceived in the extended mind theory. He argues that even in this perspective there remains a responsibility gap and concludes with a pessimistic outlook regarding the possibility to bride this gap.

Vera Borrmann, Erika Versalovic, Timothy Brown, Helena Scholl, Eran Klein, Sara Goering, Oliver Müller and Philipp Kellmeyer focus on the use of phenomenological interview methods (PIMs) to investigate the subjective experiences of individuals using neurotechnologies. To enhance and expand the methodology of phenomenological interviewing, the authors examine three different PIMs, highlighting their features and limitations. They propose a critical phenomenology approach that rejects a "neutral" subject and incorporates temporal and ecological aspects of the interviewees and interviewers, considering factors such as age, gender, social situation, bodily constitution, language skills, cognitive impairments, and traumatic memories. The authors advocate for an ethically sensitive interviewing process based on critical phenomenology and trauma-informed qualitative work, facilitating a more nuanced exploration of the interviewees' relationship with their neuro-prosthetic, and acknowledging the interpersonal and social dynamics between interviewer and interviewee.

References

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Blackwell.

Anderson, J. (2008). Neuro-Prosthetics, the extended mind, and respect for persons with disability. In M. Düwell, C. Rehmann-Sutter, & D. Mieth (Eds.), *The contingent nature of life: Bioethics and limits of human Existence* (pp. 259–274). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6764-8_22

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Springer. https://doi.org/10.1007/978-94-010-0237-0.

Churchland, P. S. (1986). *Neurophilosophy*. MIT Press.

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind (Active externalism). *Analysis, 58*(1), 7–19. https://doi.org/10.1111/1467-8284.00096.

Coninx, S. (2023). *The dark side of niche construction: Challenges in modern medicine & healthcare*. PhilSci Archive. http://philsci-archive.pitt.edu/21799/.

Craver, C. F. (2009). *Explaining the brain*. Clarendon Press.

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association, 60*(3), 441–458. https://doi.org/10.2307/3131782

Emily, B., Mathieu, G., & Racine, E. (2009). Preparing the ethical future of deep brain stimulation. *Surgical Neurology, 72*(6), 577–586.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy, LXVII, I*(1), 5–20.

Giere, R. (2002). Scientific cognition as distributed cognition. In M. Siegal, P. Carruthers, & S. Stich (Eds.), *The cognitive basis of science* (pp. 285–299). Cambridge University Press. https://doi.org/10.1017/CBO9780511613517.016.

Gilbert, F. (2015). A threat to autonomy? The intrusion of predictive brain implants. *AJOB Neuroscience, 6*(4), 4–11. https://doi.org/10.1080/21507740.2015.1076087.

Greene, J., & Cohen, J. (2006). For the law, neuroscience changes nothing and everything. In O. Goodenough & S. Zeki (Eds.), *Law and the Brain* (pp. 207–226). Oxford University Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences, 14*(3), 577–598. https://doi.org/10.1007/s11097-014-9355-1.

Heersmink, R. (2016). The cognitive integration of scientific instruments: Information, situated cognition, and scientific practice [journal article]. *Phenomenology and the Cognitive Sciences, 15*(4), 517–537. https://doi.org/10.1007/s11097-015-9432-0

Heinrichs, J.-H. (2018). Neuroethics, cognitive technologies and the extended mind perspective. *Neuroethics, 14*(1), 59–72. https://doi.org/10.1007/s12152-018-9365-8

Heinrichs, J.-H. (2021). The case for biotechnological exceptionalism. *Medicine, Health Care and Philosophy, 24*(4), 659–666. https://doi.org/10.1007/s11019-021-10032-5

Heinrichs, J.-H., Rüther, M., Stake, M., & Ihde, J. (2022). *Neuroenhancement*. Karl Alber. https://doi.org/10.5771/9783495999615

Hernández-Orallo, J., & Vold, K. (2019). *AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI* Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA. https://doi.org/10.1145/3306618.3314238

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Jebari, K. (2013). Brain Machine Interface and Human Enhancement – An Ethical Review. *Neuroethics, 6*(3), 617–625. https://doi.org/10.1007/s12152-012-9176-2

Klaming, L., & Haselager, P. (2013). Did My Brain Implant Make Me Do It? Questions Raised by DBS Regarding Psychological Continuity, Responsibility for Action and Mental Competence. *Neuroethics, 6*(3), 527–539. https://doi.org/10.1007/s12152-010-9093-1

Knorr-Cetina, K. (1999). *Epistemic cultures*. Harvard University Press.

Kraemer, F. (2013). Authenticity or autonomy? When deep brain stimulation causes a dilemma. *Journal of Medical Ethics, 39*(12), 757–760. https://doi.org/10.1136/medethics-2011-100427

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences, 23*(1), 131–146. https://doi.org/10.1017/S0140525X00002417

Levy, N. (2007). *Neuroethics*. Cambridge University Press.

Liao, S.-y., & Huebner, B. (2021). Oppressive Things*. *Philosophy and Phenomenological Research*, 103(1), 92–113. https://doi.org/10.1111/phpr.12701

Mashour, G. A., Walker, E. E., & Martuza, R. L. (2005). Psychosurgery: Past, present, and future. *Brain research. Brain research reviews, 48*(3), 409–419. https://doi.org/10.1016/j.brainresrev.2004.09.002

Nersessian, N. J. (2009). How Do Engineering Scientists Think? Model-Based Simulation in Biomedical Engineering Research Laboratories. *Topics in Cognitive Science, 1*(4), 730–757. https://doi.org/10.1111/j.1756-8765.2009.01032.x

Newen, A., Bruin, L. D., & Gallagher, S. (Eds.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.001.0001.

Nogueira de Carvalho, F., & Krueger, J. (2023). Biases in Niche Construction. *Philosophical Psychology*, 1–31.

Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.

Odling-Smee, J., & Laland, K. N. (2011). Ecological Inheritance and Cultural Inheritance: What Are They and How Do They Differ? *Biological Theory, 6*(3), 220–230. https://doi.org/10.1007/s13752-012-0030-x

Putnam, H. (1975). The meaning of 'meaning.' In K. Gunderson (Ed.), *Language, mind, and knowledge* (pp. 131–193). University of Minnesota Press.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge University Press.

Robbins, P., & Aydede, M. (Eds.). (2009). *The Cambridge handbook of situated cognition*. Cambridge University Press.

Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Science, 10*(9), 419–423. https://doi.org/10.1016/j.tics.2006.07.011

Schlaepfer, T., & Fins, J. J. (2010). Deep Brain Stimulation and the Neuroethics of Responsible Publishing: When One Is Not Enough. *JAMA, 303*(8), 775–776. https://doi.org/10.1001/jama.2010.140

Shapiro, L. (Ed.). (2014). *The Routledge handbook of embodied cognition*. Routledge/Taylor & Francis Group.

Slaby, J. (2016). Mind Invasion: Situated Affectivity and the Corporate Life Hack [Hypothesis and Theory]. *Frontiers in Psychology, 7*,. https://doi.org/10.3389/fpsyg.2016.00266

Soekadar, S. R., Haagen, K., & Birbaumer, N. (2008). Brain-Computer Interfaces (Bci): Restoration of Movement and Thought from Neuroelectric and Metabolic Brain Activity. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, Behavioral and Social Dynamics* (pp. 229–252). Springer.

Synofzik, M., & Schlaepfer, T. E. (2008). Stimulating personality: Ethical criteria for deep brain stimulation in psychiatric patients and for enhancement purposes. *Biotechnology Journal, 3*(12), 1511–1520. https://doi.org/10.1002/biot.200800187

Timms, R., & Spurrett, D. (2023). Hostile Scaffolding. *Philosophical Papers*, (online first) https://doi.org/ https://doi.org/10.1080/05568641.2023.2231652

Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology, 28*(1), 107–124. https://doi.org/10.1007/s13347-014-0156-9

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind*. MIT Press.

The Ethics of the Extended Mind: Mental Privacy, Manipulation and Agency

Robert W Clowes, Paul Smart and Richard Heersmink

1 A New Ethical Landscape?

The extended mind hypothesis has been one of the most influential ideas originating in philosophy over the last 25 years. According to this hypothesis, artefacts, objects and other individuals may count as a constitutive part of a person's mind (Clark & Chalmers 1998; Clark 2008). There has been much debate about the metaphysics of the extended mind, but, until recently, the practical and normative consequences have been little explored. This is starting to change (e.g., Levy 2007; Heersmink 2017a, b; Heinrichs 2017, 2021; Carter & Palermos 2016; Clowes 2015). The extended mind hypothesis has changed the way we think about our relation to the local environment, and ethical issues are a plausible next step in its intellectual trajectory.

For those unfamiliar with the extended mind, we will briefly introduce the case of Otto (Clark & Chalmers 1998). Otto is a man afflicted by a deterioration in bio-mnemonic capabilities, incurred as the result of a mild form of dementia. As a coping strategy, Otto uses a notebook to aid him in remembering important information. Thus, when Otto is in New York and desires to visit the Museum

R. W. Clowes
Instituto de Filosofa da Nova, Universidade Nova de Lisboa, Lisbon, Portugal
e-mail: robertclowes@fcsh.unl.pt

P. Smart
Electronics & Computer Science, University of Southampton, Southampton, United Kingdom
e-mail: ps02v@ecs.soton.ac.uk

R. Heersmink (✉)
Department of Philosophy, Tilburg University, Tilburg, Netherlands
e-mail: j.r.heersmink@tilburguniversity.edu

J.-H. Heinrichs et al. (eds.), *Neuro-ProsthEthics*, Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie 9,
https://doi.org/10.1007/978-3-662-68362-0_2

of Modern Art (MoMA), he automatically consults his notebook and retrieves the information that MoMA is located on 53rd Street. According to Clark and Chalmers, the information in Otto's notebook plays more or less the same role in guiding Otto's thoughts and actions as does the information typically stored in biological memory. Given this, Clark and Chalmers suggest that we ought to regard the notebook (and its informational contents) as part of the supervenience base of Otto's dispositional beliefs. If the information had been retrieved from bio-memory, they suggest, then we would have little problem in regarding the bio-memory system as part of the supervenience base for Otto's beliefs (and thus a *bona fide* part of the machinery of his mind). Given this, however, it is hard to see why we ought to regard the notebook any differently. If both the notebook and bio-memory provide us with a suitable folk psychological grip over Otto's overt behaviour, then perhaps they both ought to be afforded equal cognitive status. That is to say, they both ought to be regarded as *bona fide* constituents of Otto's mind.

In support of such claims, Clark and Chalmers refer to a set of criteria that have come to be known as the trust + glue criteria.[1] In short, Clark and Chalmers claim that what makes the notebook part of Otto's mind is the fact that Otto has a certain relation to the notebook. What is crucial to the Otto case, Clark and Chalmers suggest, is that Otto has a high degree of trust in the notebook, he relies upon it, and it is readily accessible. When Otto desires to go to MoMA, he automatically consults the notebook, the relevant information is easily retrieved, and, upon accessing it, Otto automatically endorses it—he does not subject it to critical scrutiny in the way that we might treat information from a suspect news source.

In this chapter, we focus on exploring some of the ethical issues associated with the extended mind. We also reflect on how the extended mind—and the broader notion of cognitive extension[2]—might help us reframe new aspects of the ethical landscape that we inhabit. The idea of the extended mind may be particularly apposite to our historical moment, and its ethical implications especially useful to follow-through. It is thus worth briefly exploring why the concept of the extended mind has been so influential in recent times. There are arguably two main reasons for this.

First, much contemporary cognitive science has a strongly anti-Cartesian orientation that emphasises the need to understand cognition in its active, world-involving situated and embodied forms. So-called 4E cognitive science—emphasising the embodied, embedded, enactive, and extended nature of cognition—has

[1] A number of criteria have been proposed to individuate cases of cognitive and mental extension. For reasons of space, we do not consider these additional criteria. See Heersmink (2015), for a review of some of the criteria that have been discussed in the literature.

[2] A distinction is sometimes made between extended cognition and the extended mind, with the former centred on explanatory kinds relevant to cognitive science (e.g., extended problem-solving), and the latter centred on explanatory kinds relevant to folk psychology (e.g., dispositional belief). In the present paper, the term "cognitive extension" should be understood as referring to both extended cognition and the extended mind.